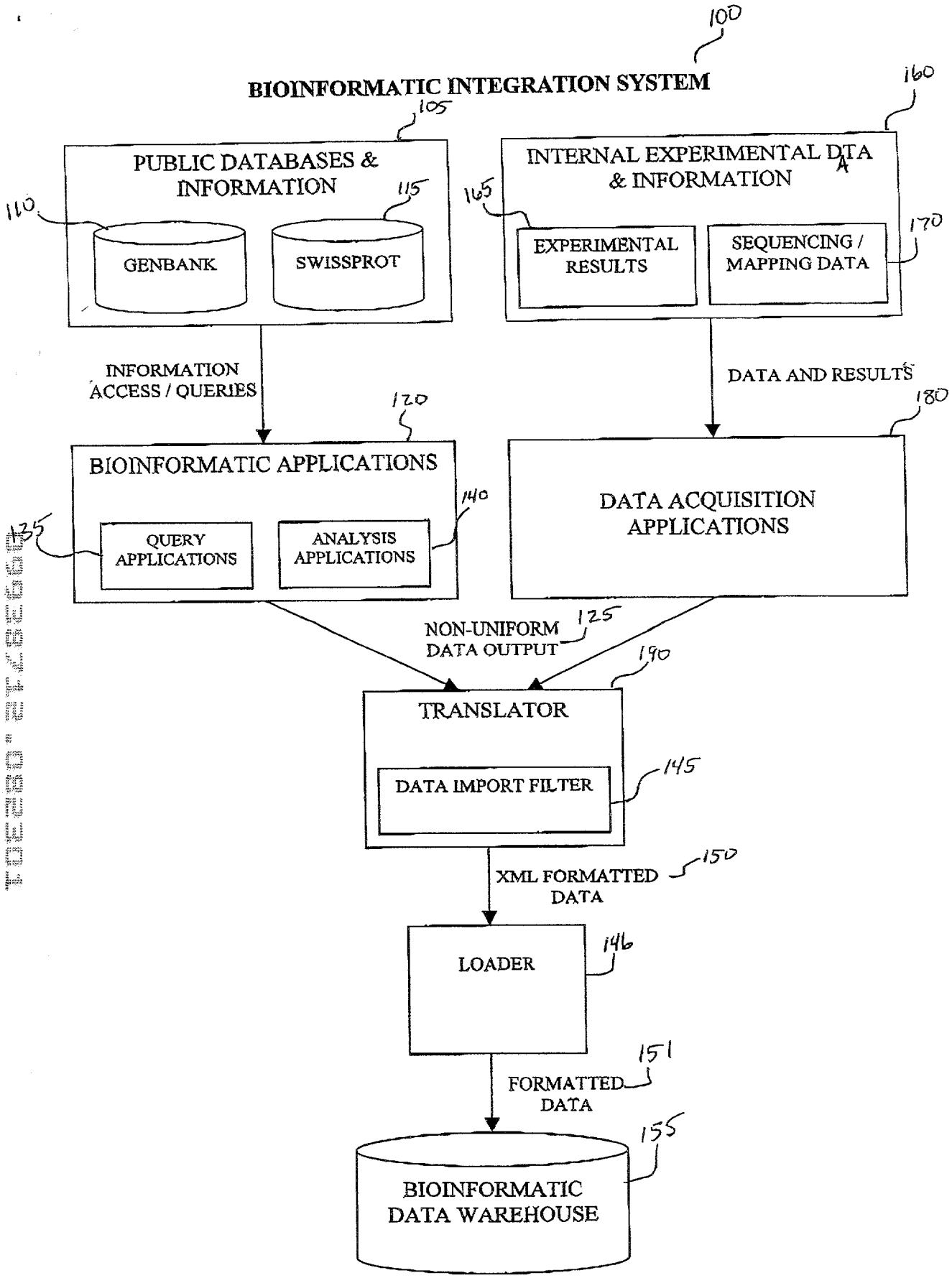
100
BIOINFORMATIC INTEGRATION SYSTEM
105
PUBLIC DATABASES & INFORMATION
110
GENBANK
115
SWISSPROT
160
INTERNAL EXPERIMENTAL DTA & INFORMATION
165
EXPERIMENTAL RESULTS
170
SEQUENCING / MAPPING DATA
INFORMATION ACCESS / QUERIES
DATA AND RESULTS
120
BIOINFORMATIC APPLICATIONS
135
QUERY APPLICATIONS
140
ANALYSIS APPLICATIONS
180
DATA ACQUISITION APPLICATIONS
NON-UNIFORM DATA OUTPUT
125
190
TRANSLATOR
145
DATA IMPORT FILTER
XML FORMATTED DATA
150
LOADER
146
151
FORMATTED DATA
155
BIOINFORMATIC DATA WAREHOUSE

FIGURE 1

LOADER FOR BIOINFORMATIC INTEGRATION SYSTEM

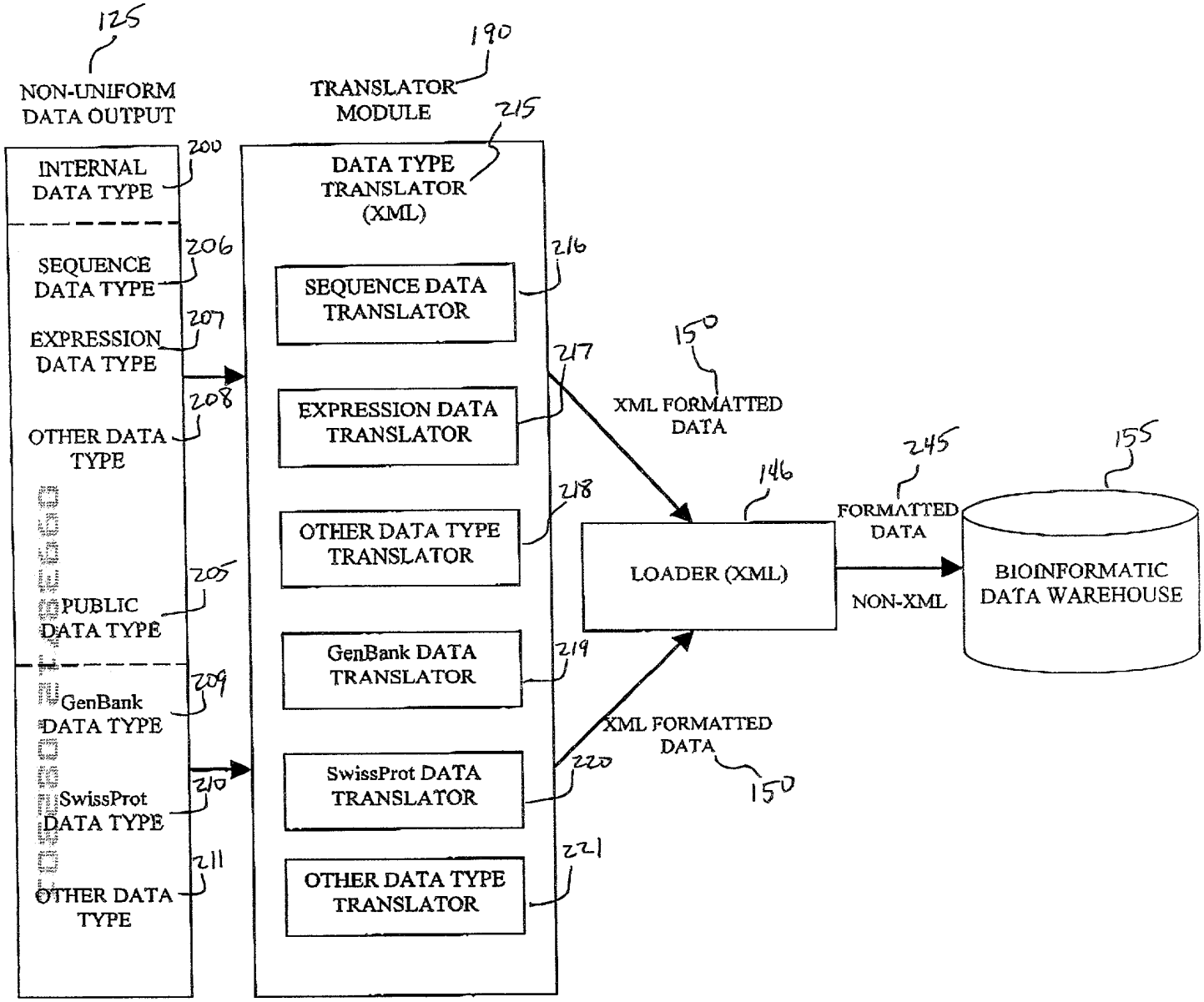

FIGURE 2

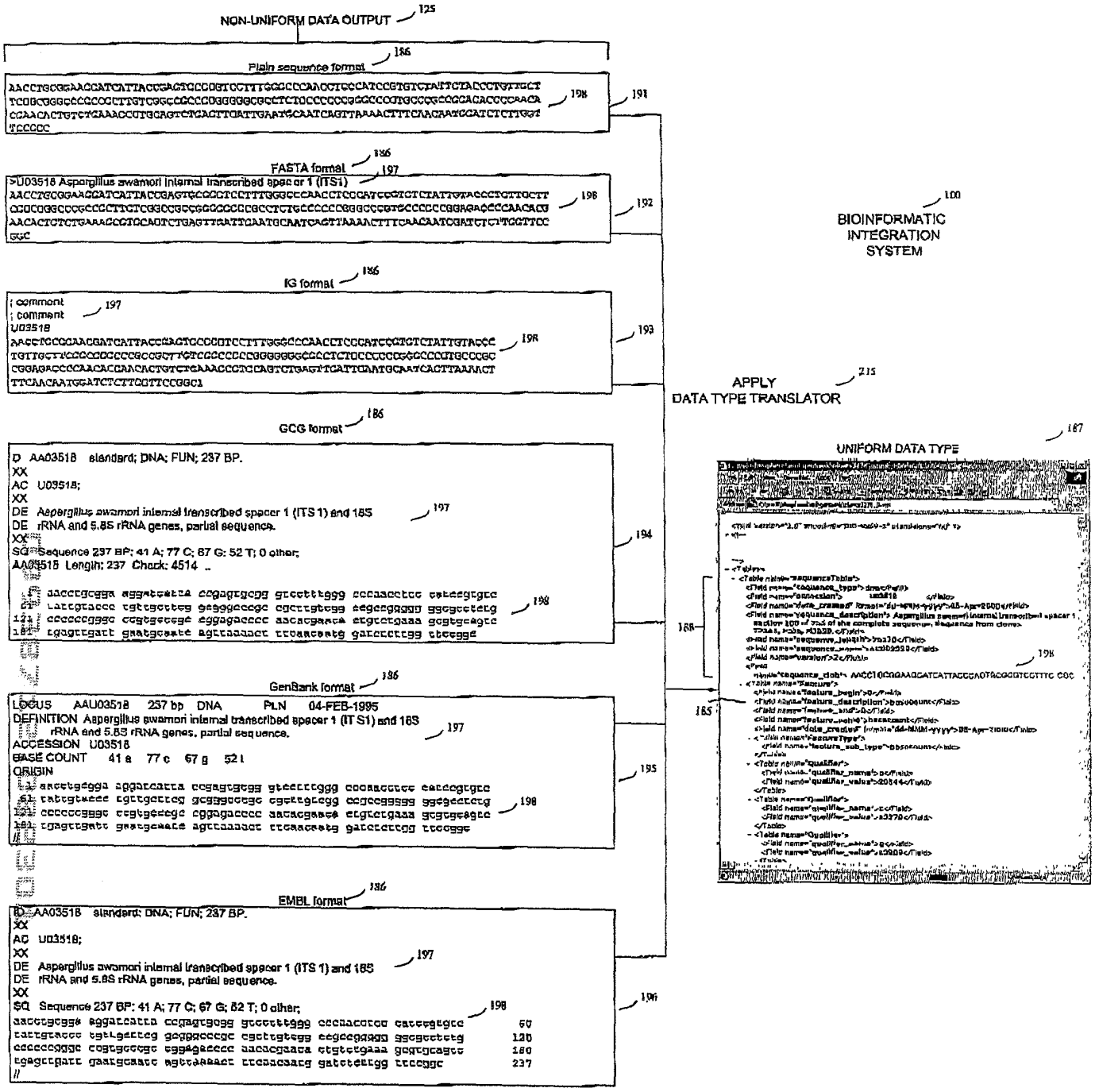
NON-UNIFORM DATA OUTPUT
125
Plain sequence format
186
198
191
FASTA format
186
197
>U03518 Aspergillus awamori internal transcribed spacer 1 (ITS1)
198
192
100
BIOINFORMATIC INTEGRATION SYSTEM
IG format
186
; comment
; comment
197
U03518
198
193
APPLY DATA TYPE TRANSLATOR
215
GCG format
186
D AA03518 standard; DNA; FUN; 237 BP.
AC U03518;
DE Aspergillus awamori internal transcribed spacer 1 (ITS 1) and 18S
DE rRNA and 5.8S rRNA genes, partial sequence.
197
SQ Sequence 237 BP; 41 A; 77 C; 67 G; 52 T; 0 other;
AA03518 Length: 237 Check: 4514 ..
198
194
UNIFORM DATA TYPE
187
188
185
GenBank format
186
LOCUS AAU03518 237 bp DNA PLN 04-FEB-1995
DEFINITION Aspergillus awamori internal transcribed spacer 1 (ITS1) and 18S
rRNA and 5.8S rRNA genes, partial sequence.
197
ACCESSION U03518
BASE COUNT 41 a 77 c 67 g 52 t
ORIGIN
198
195
EMBL format
186
ID AA03518 standard; DNA; FUN; 237 BP.
AC U03518;
DE Aspergillus awamori internal transcribed spacer 1 (ITS 1) and 18S
DE rRNA and 5.8S rRNA genes, partial sequence.
197
SQ Sequence 237 BP; 41 A; 77 C; 67 G; 52 T; 0 other;
198
196

FIGURE 3

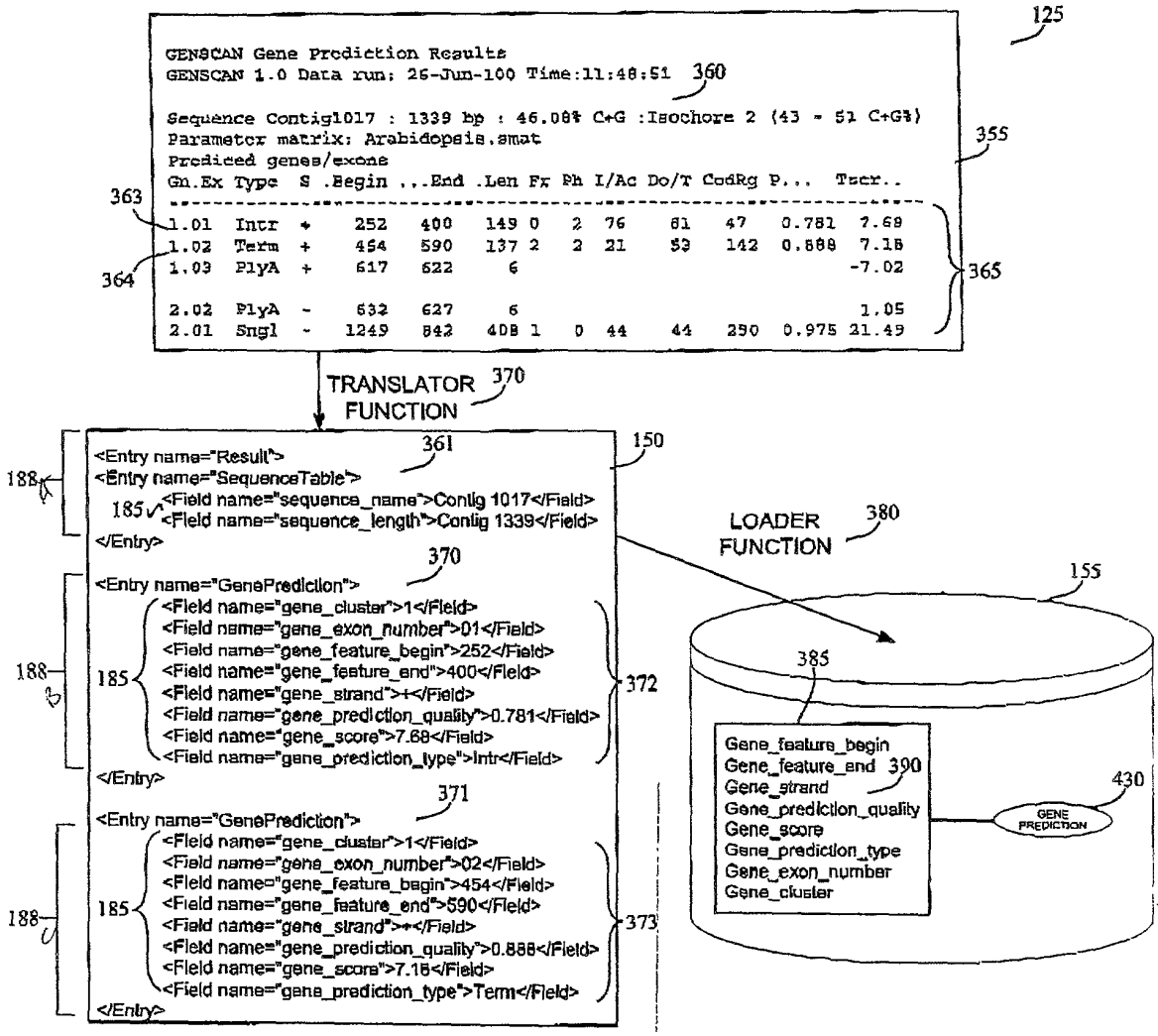
125
GENSCAN Gene Prediction Results
GENSCAN 1.0 Data run: 26-Jun-100 Time:11:48:51
360
Sequence Contig1017 : 1339 bp : 46.08% C+G :Isochore 2 (43 - 51 C+G%)
Parameter matrix: Arabidopsis.smat
Predicted genes/exons
355
Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P... Tscr..
363
1.01 Intr + 252 400 149 0 2 76 61 47 0.781 7.68
1.02 Term + 454 590 137 2 2 21 59 142 0.888 7.16
364
1.03 PlyA + 617 622 6 -7.02
365
2.02 PlyA - 632 627 6 1.05
2.01 Sngl - 1249 842 408 1 0 44 44 290 0.975 21.49
TRANSLATOR FUNCTION
370
150
361
188
<Entry name="Result">
<Entry name="SequenceTable">
185
<Field name="sequence_name">Contig 1017</Field>
<Field name="sequence_length">Contig 1339</Field>
</Entry>
370
<Entry name="GenePrediction">
188
185
<Field name="gene_cluster">1</Field>
<Field name="gene_exon_number">01</Field>
<Field name="gene_feature_begin">252</Field>
<Field name="gene_feature_end">400</Field>
<Field name="gene_strand">+</Field>
<Field name="gene_prediction_quality">0.781</Field>
<Field name="gene_score">7.68</Field>
<Field name="gene_prediction_type">Intr</Field>
372
</Entry>
371
<Entry name="GenePrediction">
188
185
<Field name="gene_cluster">1</Field>
<Field name="gene_exon_number">02</Field>
<Field name="gene_feature_begin">454</Field>
<Field name="gene_feature_end">590</Field>
<Field name="gene_strand">+</Field>
<Field name="gene_prediction_quality">0.888</Field>
<Field name="gene_score">7.16</Field>
<Field name="gene_prediction_type">Term</Field>
373
</Entry>
LOADER FUNCTION
380
155
385
Gene_feature_begin
Gene_feature_end
390
Gene_strand
Gene_prediction_quality
Gene_score
Gene_prediction_type
Gene_exon_number
Gene_cluster
430
GENE PREDICTION

FIGURE 4

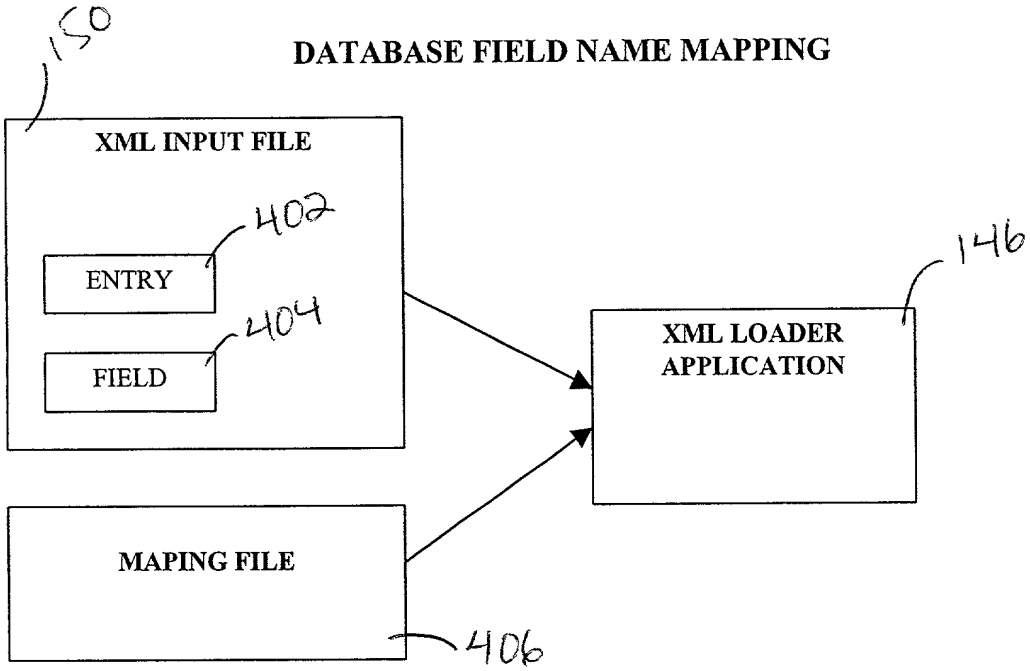
150
DATABASE FIELD NAME MAPPING
XML INPUT FILE
402
ENTRY
404
FIELD
146
XML LOADER APPLICATION
MAPING FILE
406

FIGURE 5

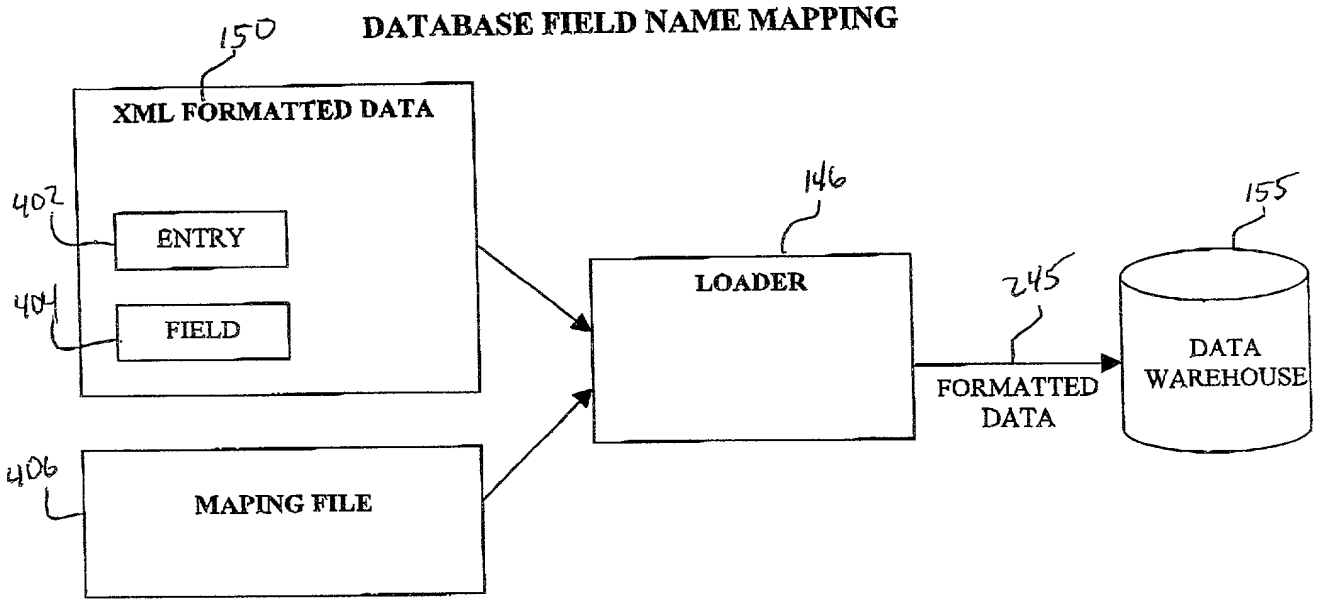

FIGURE 5

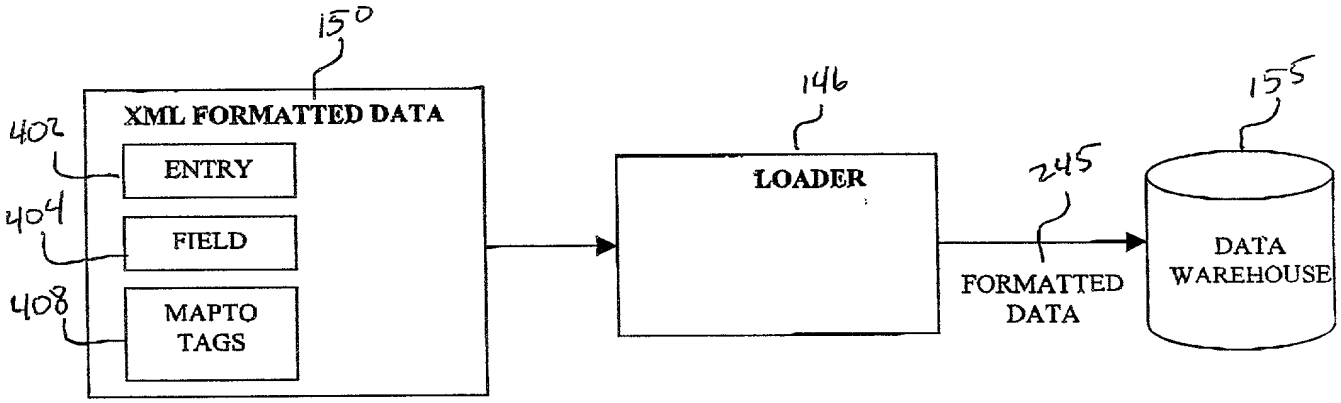

FIGURE 6

DATA MAPPING MODULE 500

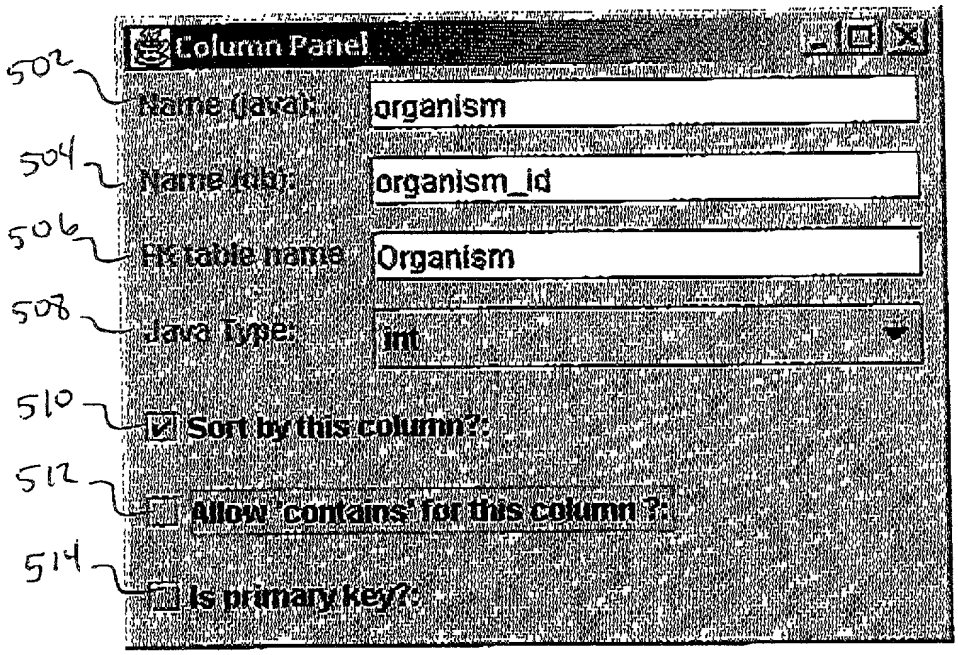

FIGURE 7

LOADER FUNCTIONALITY

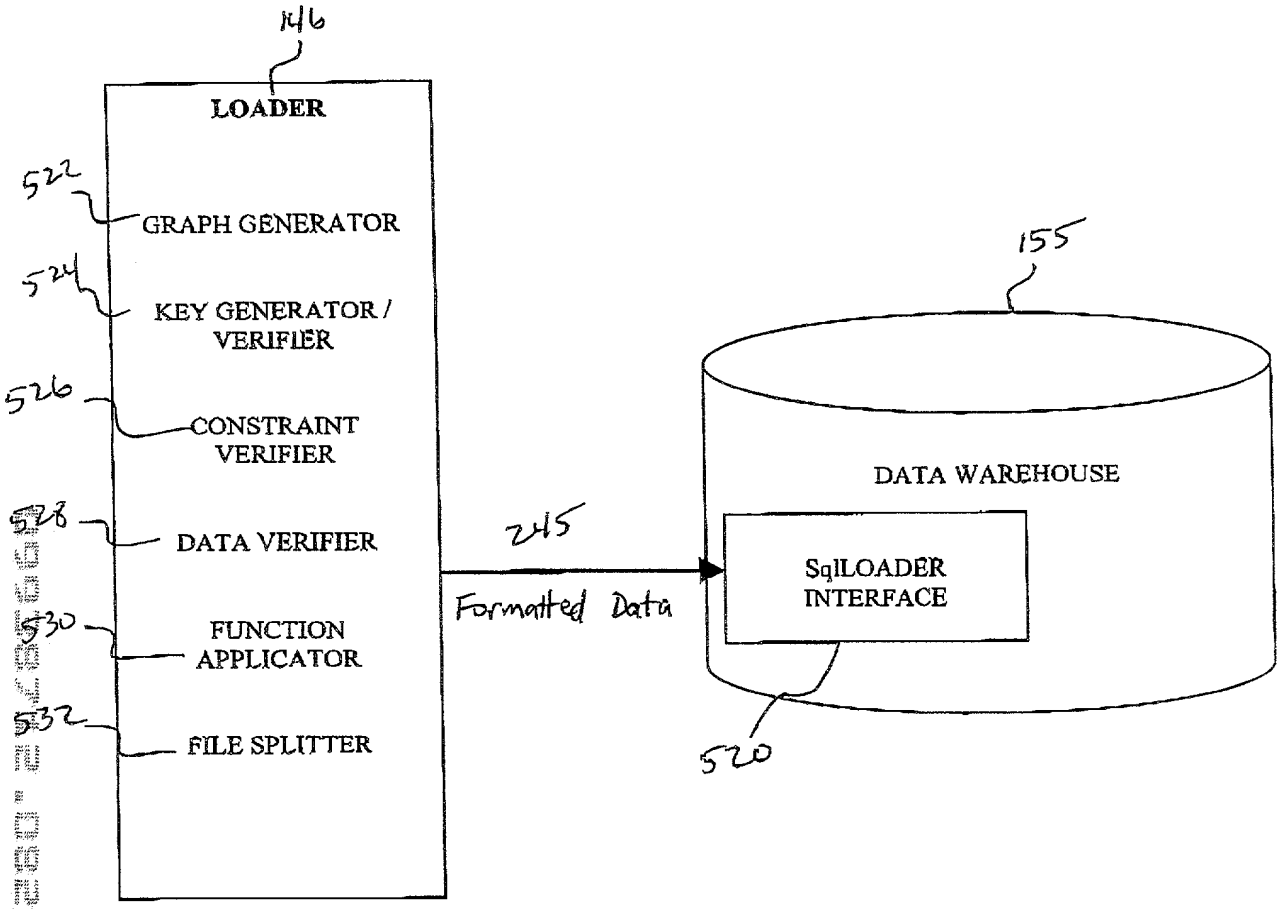

FIGURE 8

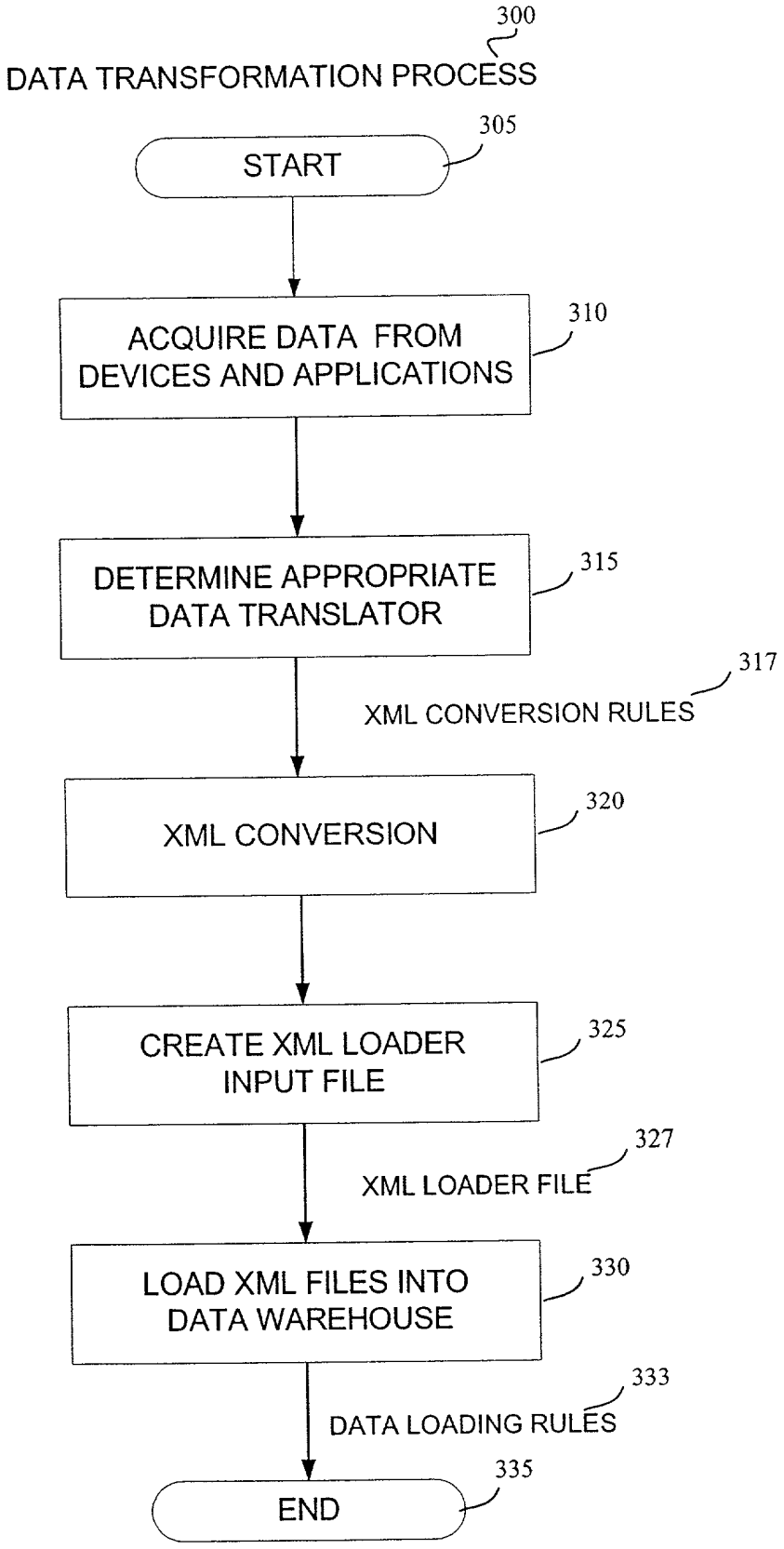
300
DATA TRANSFORMATION PROCESS
305
START
310
ACQUIRE DATA FROM DEVICES AND APPLICATIONS
315
DETERMINE APPROPRIATE DATA TRANSLATOR
317
XML CONVERSION RULES
320
XML CONVERSION
325
CREATE XML LOADER INPUT FILE
327
XML LOADER FILE
330
LOAD XML FILES INTO DATA WAREHOUSE
333
DATA LOADING RULES
335
END

FIGURE 9